

Data Release Statement **GRID3 Zambia Settlement Extents, Version 01**

July 2021

Abstract

This document outlines the methodological approach and data sources used to construct the “GRID3 Zambia Settlement Extents, Version 01 dataset.” The dataset consists of settlement extents across Zambia , as well as accompanying population estimates for each settlement extent. Limitations and use constraints are provided.

Dataset citation

Center for International Earth Science Information Network (CIESIN), Columbia University and Novel-T. 2021. GRID3 Zambia Settlement Extents, Version 01. Palisades, NY: Geo-Referenced Infrastructure and Demographic Data for Development (GRID3). <https://doi.org/10.7916/d8-vc0a-nw10> .Accessed DAY MONTH YEAR.

Data use constraints

Users are free to use, copy, distribute, transmit, and adapt the work for commercial and non-commercial purposes, without restriction, as long as clear attribution of the source is provided.

Copyright 2021. The Trustees of Columbia University in the City of New York.

The data and accompanying document is licensed under a Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0>) and specified in legal code (<http://creativecommons.org/licenses/by/4.0/legalcode>).

Contacts and data queries

GRID3 appreciates feedback regarding this dataset, such as suggestions, discovery of errors, difficulties in using the data, and format preferences.

For dataset-related questions, please send an email to any of the following organisations:

GRID3: data.queries@grid3.org

Table of Contents

Introduction	3
Methodological approach	3
Input data	3
Methods	4
Dataset description	7
Dataset codebook	7
Version History	9
Known data limitations and disclaimer	9
Acknowledgments	9
References	10

Introduction

Settlement extents are polygons representing areas where there is likely a human settlement based on the presence of buildings detected in satellite imagery. Settlement extents are not meant to represent the boundaries of an administrative unit or locality. A single settlement extent may be made up of multiple localities, especially in urban areas. Each settlement extent has an associated population estimate. This document details the methodology applied to produce the dataset “GRID3 Zambia Settlement Extents, Version 01.”

This work has been undertaken as part of the Geo-Referenced Infrastructure and Demographic Data for Development (GRID3) programme. The programme is funded by the Bill & Melinda Gates Foundation and the United Kingdom's Foreign, Commonwealth & Development Office. It is implemented by the Flowminder Foundation, WorldPop at the University of Southampton, the United Nations Population Fund, and the Center for International Earth Science Information Network (CIESIN) at Columbia University.

Methodological approach

Input data

Settlement extents and associated population estimates are derived from two primary input datasets: building footprints and population estimates (**Table 1**).

Table 1. Input datasets

<i>Data</i>	<i>Description</i>	<i>Source</i>
Building footprints	Building footprints are derived by the Ecopia Tech Corporation using a feature extraction algorithm. The dataset consists of polygon data delineating rooftops present in satellite imagery captured between 2009 and 2019.	Maxar Technologies, Inc. and Ecopia Tech Corporation. 2020. Ecopia Landbase Africa powered by Maxar. Available at DigitizeAfrica.ai
WorldPop Population estimates	<u>Population estimates (2020)</u> : Top-down modeled population estimates (100-meter resolution) produced by WorldPop at the University of	Bondarenko M., Kerr D., Sorichetta A., and Tatem, A.J. 2020. Census/projection-

	<p>Southampton. Units: Total population per pixel. Extent: Zambia https://www.worldpop.org/methods/populations</p> <p><u>UN-adjusted Population estimates (2020):</u> Identical to above, but country population totals for Zambia are adjusted to match estimates prepared by: United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019, Online Edition. Rev. 1.</p>	<p>disaggregated gridded population datasets for 51 countries across sub-Saharan Africa in 2020 using building footprints. WorldPop, University of Southampton, UK. doi:10.5258/SOTON/WP00682</p> <p>doi:10.5258/SOTON/WP00683 (UN-adjusted)</p>
WorldPop GRID3 Gridded Population Estimates	<p><u>Zambia gridded population estimates:</u> Bottom-up modeled estimates of total population size for each 100m grid cell. https://www.worldpop.org/methods/populations</p> <p><u>Zambiagridded population estimates for specific age-sex groups:</u> Bottom-up modeled estimates of population sizes for individual age-sex groups for each 100m grid cell.</p>	<p>WorldPop and Institut National de la Statistique et de la Demographie du Zambia. 2020. Census-based gridded population estimates for Zambia (2019), version 1.0. WorldPop, University of Southampton. doi:10.5258/SOTON/WP00687</p>

Methods

Both the settlements' extents and classification are derived solely from Ecopia's building footprints. Data analysis and processing were achieved entirely using Esri's ArcGIS software (Pro version 2.7.3), it's native module arcpy, and open source python library pandas 1.3.0.

The center points of building footprint features are converted to a raster grid (3 arc-second resolution) of building densities (i.e. total number of building points within each cell), referred to throughout this document as "building densities." The ArcGIS Pro "Contour" tool was used to generate shell-up contours from the building density grid



to delineate settled versus non-settled areas. The shell-up method includes contours that start at the lower bounds, but includes all grid cells with building densities to the upper bounds of the grid. For example, a shell-up contour of 10 would include all grid cells with a building density of 10 or more. Contours with a building density of one or more are used to create the settlement extent polygons (see **Figure 1**).

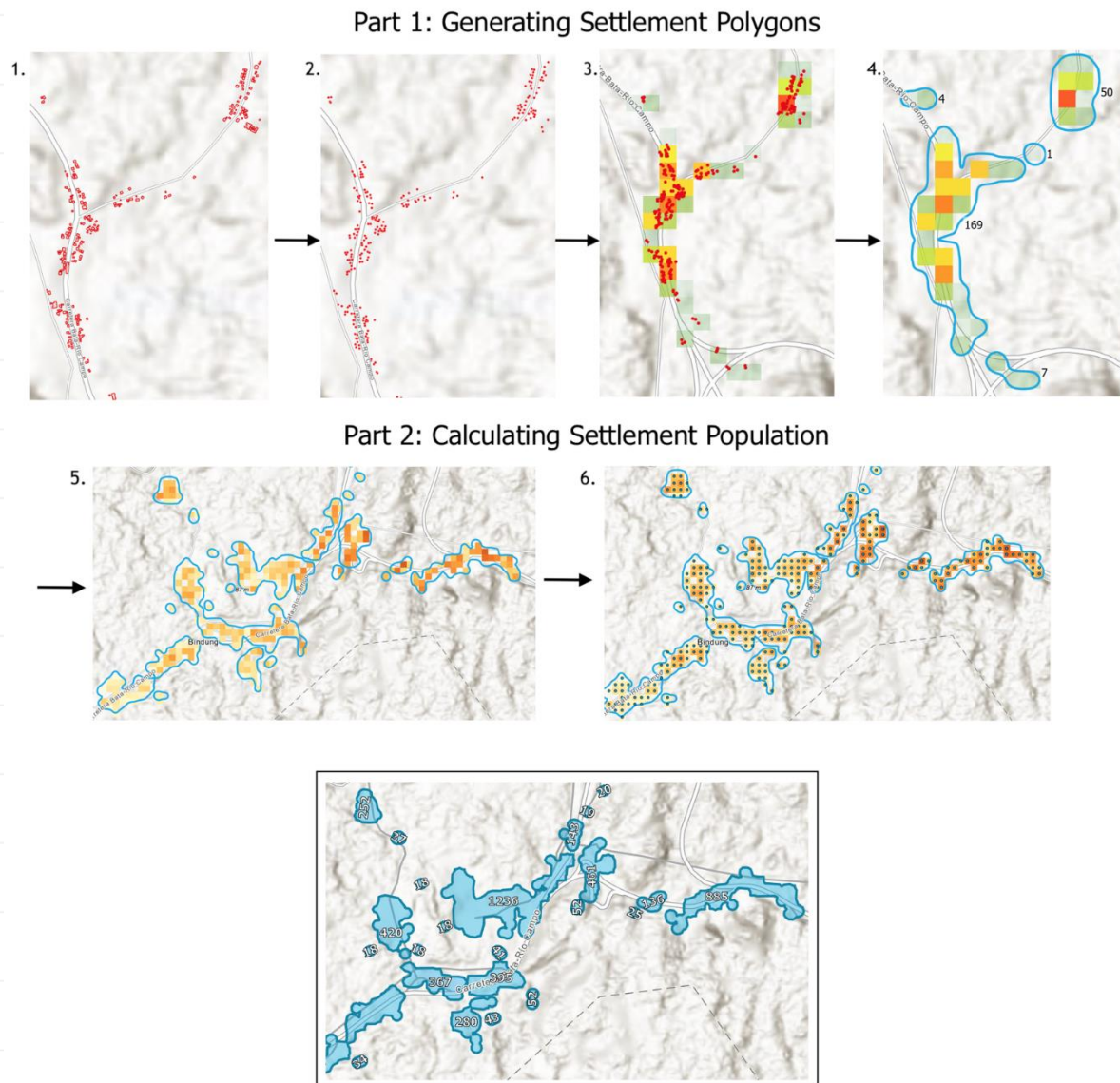


Figure 1: Part 1: Illustration of methods used to generate settlement extents: 1) Building footprints (Ecopia/Maxar) are projected to map and 2) Building footprints are converted to points. 3) Building points are converted to a building density raster layer at 100m resolution. Values represent the total number of building points within each cell. 4) Shell-up contours are generated around building density cells, and building counts within each settlement extent are summed. Contour geometry is smoothed to produce final settlement extent polygons. **Part 2:** Illustration of methods to generate

population estimates for each settlement extent: 5) A population raster (100m, WorldPop) is added to the map and 6) cells are converted to points. Each population point is assigned to the nearest settlement extent and population values are summed to produce population estimates. Steps 5 and 6 are performed once for a WorldPop raster and once for a UN-adjusted WorldPop raster. **Bottom panel** represents final settlement extents with associated population count (rounded to nearest integer for clarity).

The GRID3 Zambia settlement extents are then classified (based on building density) into three classes: built-up areas (BUA), small settlement areas (SSA), and Hamlets (Barau et al., 2014).

Table 2. Settlement extent classification

Settlement class	Description
Built-up areas (BUAs)	A built-up area (BUA) is generally an area of urbanisation with moderately-to-densely-spaced buildings and a visible grid of streets and blocks. BUAs are characterised as polygons that maintain a 100 m ² building density of 13 or more across an area greater than or equal to 0.4 km ² .
Small Settlements (SSAs)	A small settlement (SSA) is a settled area of permanently inhabited structures and compounds of roughly a few hundred to a few thousand inhabitants. The housing pattern in SSAs is an assemblage of family compounds adjoining other similar habitations. Small settlement areas are characterised as polygons containing 50 or more buildings and not a BUA.
Hamlet	A hamlet is a collection of several compounds or sleeping houses in isolation from small settlements or urban areas. Hamlets are characterised as polygons containing between 1 and 49 buildings.

The population estimates for each settlement extent were calculated using population estimates from WorldPop. Individual, country-constrained population raster cells and their associated population estimate (100m resolution) were converted to individual points and joined to the nearest settlement extent. Finally, the values within a settlement extent were summed to obtain the total population for each settlement extent. Settlement extents from population estimates can be found in the “Population” and “Pop_UN_adj” fields.

Dataset description

The data are in geodatabase format and consist of a single-feature class. An example of data is shown on Figure 3.

Extent: Zambia: Admin Level 0 Boundaries. The overall extent of the layer is limited to the overall extent of the building footprint dataset and may not reflect the extent of official administrative boundaries.

Coordinate system: GCS WGS 1984

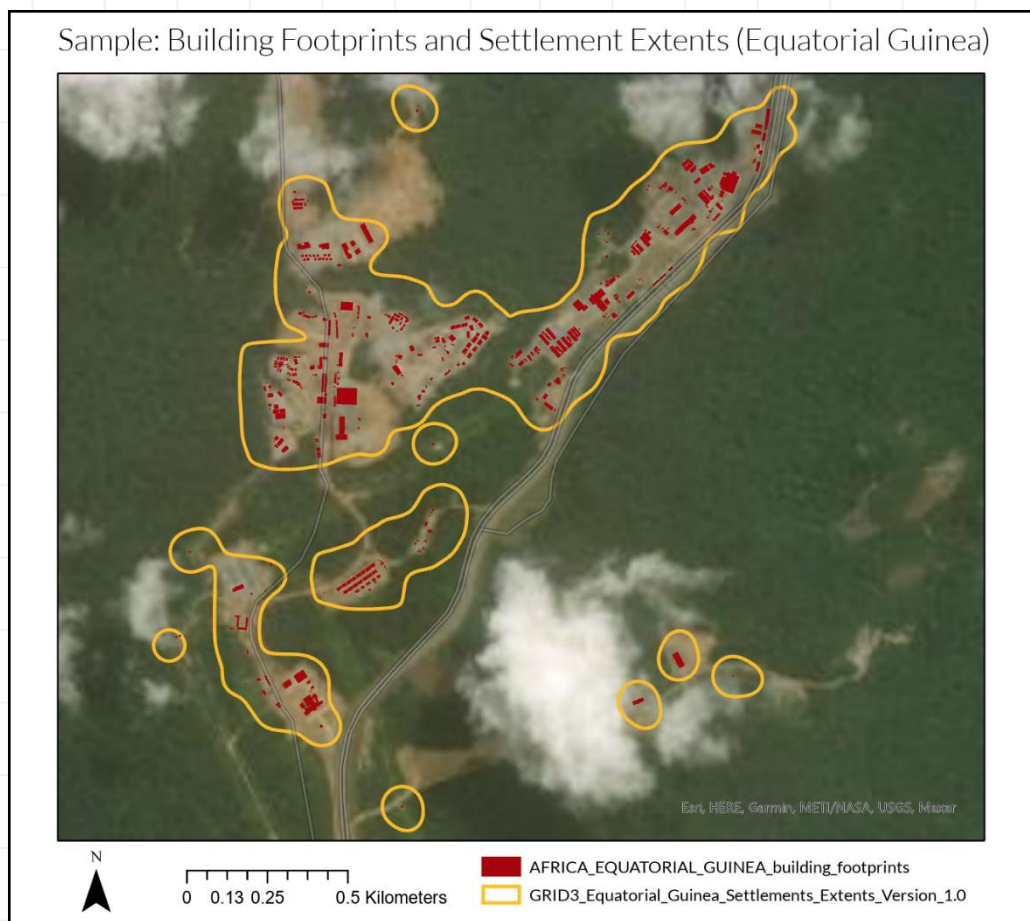


Figure 3: Sample map depicting a settlement extent data layer, with building footprint layer for reference. Note: Building footprint layers are not included in this data product.

Dataset codebook

Field	Description
OBJECTID	Default ESRI field: Unique sequential numeric identifier maintained by the database.
Shape	Default ESRI field: Geometry type
MGRS_Code	Unique name generated using the Military Grid Reference System
Country	Country name
ISO	Three-letter country ISO code
Type	The settlement type as defined by a built-up area, small-settlement area, or hamlet
Population	Total population estimate for settlement extent derived from constrained WorldPop raster
Pop_UN_adj	Total population estimate for settlement extent derived from constrained and UN-adjusted WorldPop raster
Pop_WP_GRID3	Total population estimate for settlement extent derived from WorldPop GRID3 bottom-up gridded population estimates (v1.0).
Pop_WP_GRID3_uncertainty	Quantified uncertainty of the Total population for settlement extent derived from WorldPop GRID3 bottom-up gridded population estimates (v1.0).
m0, m1, m5... f0, f1, f5...f80	Population estimates (bottom-up) for specific age-sex groups (WorldPop GRID3 bottom-up model v1.0): “f0” and “m0” are population counts of under 1-year olds for females and males, respectively. “f1” and “m1” are population counts of 1 to 4 year olds for females and males, respectively. Over 4 years old, the age groups are in five year bins labelled with a “5”, “10”, etc. Eighty year olds and over are represented in the groups “f80” and “m80.”
under1, under5, under15, f15_49	Population estimates (bottom-up) for particular demographics (GRID3 bottom-up model v1.0): “under1” (all females and males under the age of 1), “under5” (all females and males under the age

	of 5), “under15” (all females and males under the age of 15) and “f15_49” (all females between the ages of 15 and 49, inclusive).
Shape_Length	Default ESRI field: The shape length in geographic coordinates (decimal degrees)
Shape_Area	Default ESRI field: The shape area in geographic coordinates (decimal degrees)

Version History

This data product contains all information contained in the previous “GRID3 Zambia Settlement Extents, Version 01 Alpha” product, with updates. Updates in this version include: a single settlement extent feature class (alpha version contains the same data in three separate feature class layers: BUAs, SSAs, and hamlets) and new population estimate fields (Population and Pop_UN_adj) for each settlement extent.

Known data limitations and disclaimer

The “GRID3 Zambia Settlement Extents, Version 01” has not been visually inspected for false negatives. A settlement may exist for a location where there is no building footprint data and consequently no settlement extent. This may lead to settlements not being identified. Likewise, the GRID3 settlement extents have not been visually inspected for false positives. A settlement polygon may have been falsely identified as an actual settlement. The type or use of buildings (e.g. residential, non-residential, mixed use, shed, etc) are not known.

CIESIN, Columbia University, and GRID3 follow procedures designed to ensure that data disseminated by project are of reasonable quality. If, despite these procedures, users encounter apparent errors or misstatements in the data, they should contact GRID3 at data.queries@grid3.org. CIESIN, Columbia University, and their sponsors do not guarantee the accuracy, reliability, or completeness of any data provided. We provide these data without warranty of any kind whatsoever, either expressed or implied, and shall not be liable for incidental, consequential, or special damages arising out of the use of any data provided.

Acknowledgments

Funding for the development and dissemination of this dataset was provided by the Bill & Melinda Gates Foundation and the United Kingdom's Foreign, Commonwealth & Development Office.

References

Barau, I., et al. 2014. "Improving Polio Vaccination Coverage in Nigeria Through the Use of Geographic Information System Technology, The Journal of Infectious Diseases, Volume 210, Issue suppl_1, Pages S102–S110, <https://doi.org/10.1093/infdis/jiu010>

Bondarenko M., Kerr D., Sorichetta A., and Tatem, A.J. 2020. "Census/projection-disaggregated gridded population datasets for 51 countries across sub-Saharan Africa in 2020 using building footprints." WorldPop, University of Southampton, UK. doi:10.5258/SOTON/WP00682; doi:10.5258/SOTON/WP00683

Esri Inc. 2020. ArcGIS Pro (Version 2.7.3). Esri Inc. <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>

Maxar Technologies, Inc. and Ecopia Tech Corporation. 2020. Ecopia Landbase Africa powered by Maxar, website - "DigitizeAfrica.ai".

McKinney, Wes. 2010. Data Structures for Statistical Computing in Python (Pandas), Proceedings of the 9th Python in Science Conference, 51-56